

OIKE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,538

DATE: 06/12/2001

TIME: 15:13:10

Input Set : A:\REGEN1530-2.ST25.txt

Output Set: N:\CRF3\06122001\I866538.raw

3 <110> APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA  
 4       TSIEN, Roger  
 5       Campbell, Robert  
 7 <120> TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS  
 9 <130> FILE REFERENCE: REGEN1530-2  
 C--> 11 <140> CURRENT APPLICATION NUMBER: US/09/866,538  
 12 <141> CURRENT FILING DATE: 2001-05-24  
 14 <160> NUMBER OF SEQ ID NOS: 29  
 16 <170> SOFTWARE: PatentIn version 3.0  
 18 <210> SEQ ID NO: 1  
 19 <211> LENGTH: 716  
 20 <212> TYPE: DNA  
 21 <213> ORGANISM: Aequorea victoria  
 23 <220> FEATURE:  
 24 <221> NAME/KEY: CDS  
 25 <222> LOCATION: (1)..(714)  
 27 <400> SEQUENCE: 1  
 28 atg agt aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt 48  
 29 Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val  
 30 1 5 10 15  
 32 gaa tta gat ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag 96  
 33 Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu  
 34 20 25 30  
 36 ggt gaa ggt gat gca aca tac gga aaa ctt acc ctt aaa ttt att tgc 144  
 37 Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys  
 38 35 40 45  
 40 act act gga aaa cta cct gtt cca tgg cca aca ctt gtc act act ttc 192  
 41 Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe  
 42 50 55 60  
 44 tct tat ggt gtt caa tgc ttt tca aqa tac cca gat cat atg aaa cag 240  
 45 Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln  
 46 65 70 75 80  
 48 cat gac ttt ttc aag agt gcc atg ccc gaa ggt tat gta cag gaa aqa 288  
 49 His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg  
 50 85 90 95  
 52 act ata ttt ttc aaa gat gac ggg aac tac aag aca cgt gct gaa gtc 336  
 53 Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val  
 54 100 105 110  
 56 aag ttt gaa ggt gat acc ctt gtt aat aga atc gag tta aaa ggt att 384  
 57 Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile  
 58 115 120 125  
 60 gat ttt aaa gaa gat gga aac att ctt gga cac aaa ttg gaa tac aac 432  
 61 Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn  
 62 130 135 140  
 64 tat aac tca cac aat gta tac atc atg gca gac aaa caa aag aat gga 480  
 65 Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly  
 66 145 150 155 160

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,538

DATE: 06/12/2001

TIME: 15:13:10

Input Set: A:\REGEN1530-2.ST25.txt

Output Set: N:\CRF3\06122001\I866538.raw

```

68 atc aaa gtt aac ttc aaa att aga cac aac att gaa gat gga agc gtt      528
69 ile lys val asn phe lys ile arg his asn ile glu asp gly ser val
70          165          170          175
72 gaa cta gca gac cat tat caa caa aat act cca att gac gat ggc ctt      576
73 gln leu ala asp his tyr gln gln asn thr pro ile gly asp gly pro
74          180          185          190
76 gtc ctt tta cca gac aac cat tac ctg tcc aca caa tct gcc ctt tga      624
77 val leu leu pro asp asn his tyr leu ser thr gln ser ala leu ser
78          195          200          205
80 aaa gat ccc aac gaa aag aga gac cac atg gtc ctt ctt gag ttt gta      672
81 lys asp pro asn glu lys arg asp his met val leu leu glu phe val
82          210          215          220
84 aca gct gct gag att aca cat gcc atg gat gaa cta tac aaa ta      716
85 thr ala ala gly ile thr his gly met asp glu leu tyr lys
86 225          230          235
89 <210> SEQ ID NO: 2
90 <211> LENGTH: 238
91 <212> TYPE: PRF
92 <213> ORGANISM: Aequorea victoria
94 <400> SEQUENCE: 2
96 Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val
97 1          5          10          15
100 Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu
101          20          25          30
104 Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys
105          35          40          45
108 Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Phe
109          50          55          60
112 Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln
113 65          70          75          80
116 His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg
117          85          90          95
120 Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val
121          100          105          110
124 Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile
125          115          120          125
128 Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn
129          130          135          140
132 Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly
133 145          150          155          160
136 Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val
137          165          170          175
140 Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro
141          180          185          190
144 Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser
145          195          200          205
148 Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val
149          210          215          220
152 Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,538

DATE: 06/12/2001

TIME: 15:13:10

Input Set : A:\REGEN1530-2.ST25.txt

Output Set: N:\CRF3\06122001\I866538.raw

```

153 215          230          235
156 <210> SEQ ID NO : 3
157 <211> LENGTH: 720
158 <212> TYPE: DNA
159 <213> ORGANISM: Aequorea victoria
161 <220> FEATURE:
162 <221> NAME/KEY: CDS
163 <222> LOCATION (1)..(720)
165 <400> SEQUENCE 3
166 atg gta acc aag ggc gag gag ctg ttc acc ggc gta gta ccc atc ctg      48
167 Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
168 1          5          10          15
169 ttc gag ctg gac ggc gac gta aac ggc cac aag ttc agc gta tcc ggc      96
170 Val Gln Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
171 20          25          30
172 gag ggc gaa ggc gat gcc acc tac ggc aag ctg acc ctg aag ttc atc      144
173 Glu Gly Gln Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
174 35          40          45
175 tgc acc acc ggc aag ctg ccc gta ccc taa ccc acc ctg gta acc acc      192
176 Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
177 50          55          60
178 ctg acc tac ggc gta cag tgc ttc agc cgc tac ccc gac cac atg aag      240
179 Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
180 65          70          75          80
181 cag cac gac ttc ttc aag tcc gcc atg ccc gaa ggc tac gtc cag gag      288
182 Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
183 85          90          95
184 cgc acc atc ttc ttc aag gac gac ggc aac tac aag acc cgc gcc gag      336
185 Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
186 100         105         110
187 gta aag ttc gag ggc gac acc ctg gta aac cgc atc gag ctg aag ggc      384
188 Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
189 115         120         125
190 atc gac ttc aag gag gac ggc aac atc ctg ggc cac aag ctg gag tac      432
191 Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
192 130         135         140
193 aac tac aac agc cac aac gtc tat atc atg gcc gac aag cag aag aac      480
194 Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
195 145         150         155         160
196 ggc atc aag gta aac ttc aag atc cgc cac aac atc gag gac ggc agc      528
197 Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
198 165         170         175
199 gta cag ctg gcc gac cac tac caa cag aac acc ccc atc ggc gac ggc      576
200 Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
201 180         185         190
202 ccc gta ctg ctg ccc gac aac cac tac ctg agc acc cag tcc gcc ctg      624
203 Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
204 195         200         205
205 agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg ctg gag ttc      672

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,538

DATE: 06/12/2001

TIME: 15:13:10

Input Set : A:\REGEN1530-2.SI25.txt

Output Set: N:\CRF3\06122001\I866538.raw

```

219 Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
220      210      215      220
222 atg acc gcc gcc ggg atc act ctg ggc atg gac gag ctg tac aag taa      720
223 Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Gln Leu Tyr Lys
224 225      230      235
227 <210> SEQ ID NO: 4
228 <211> LENGTH 239
229 <212> TYPE: PRT
230 <213> ORGANISM: Aequorea victoria
232 <400> SEQUENCE: 4
234 Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
235 1      5      10      15
238 Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
239      20      25      30
242 Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
243      35      40      45
246 Lys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
247      50      55      60
250 Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
251 65      70      75      80
254 Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
255      85      90      95
258 Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
259      100      105      110
262 Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
263      115      120      125
266 Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
267      130      135      140
270 Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
271 145      150      155      160
274 Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
275      165      170      175
278 Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
279      180      185      190
282 Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
283      195      200      205
286 Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
287      210      215      220
290 Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
291 225      230      235
294 <210> SEQ ID NO: 5
295 <211> LENGTH: 720
296 <212> TYPE: DNA
297 <213> ORGANISM: Aequorea victoria
299 <220> FEATURE:
300 <221> NAME/KEY: CDS
301 <222> LOCATION: (1).. (720)
303 <400> SEQUENCE: 5
304 atg ctg agc aag ggc gag gag ctg ttc acc ggg gtg gtg ccc atc ctg      48

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,538

DATE: 06/12/2001

TIME: 15:13:10

Input Set : A:\REGEN1530-2.ST25.txt

Output Set : N:\CRF3\06122001\I866538.raw

```

305 Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
306 1 5 10 15
308 atc aaq ctg gac ggc gac gta aac ggc cac agg ttc aac gtc tcc ggc 96
309 Val Glu Leu Asp Gly Asp Val Asn Gly His Arg Phe Ser Val Ser Gly
310 20 25 30
312 gag ggc gag ggc gat gcc acc tac ggc aag ctg acc ctg aag ttc atc 144
313 Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
314 35 40 45
316 tgc acc acc ggc aag ctg ccc gtc ccc tgg ccc acc ctg gtc acc acc 192
317 Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
318 50 55 60
320 ctg acc tgg ggc gtc cag tgc ttc aac cgc tac ccc gac cac atg aag 240
321 Leu Thr Trp Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
322 65 70 75 80
324 cag cac gag ttc ttc aag tcc gcc atg ccc gaa ggc tac atc cag gag 288
325 Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
326 85 90 95
328 cgt acc atc ttc ttc aac gac gac ggc aac tac aag acc cgc gcc gag 336
329 Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
330 100 105 110
332 gta aag ttc gag ggc gac acc ctg gtc aac cgc atc gag ctg aag ggc 384
333 Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
334 115 120 125
336 atc gac ttc aag gag gac ggc aac atc ctg ggg cac aag ctg gag tac 432
337 Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
338 130 135 140
340 aac tac atc agc cac aac gtc tat atc acc gcc gac aag cag aag aac 480
341 Asn Tyr Ile Ser His Asn Val Tyr Ile Thr Ala Asp Lys Gln Lys Asn
342 145 150 155 160
344 ggc atc aag gcc cac ttc aag atc cgc cac aac atc gag gac ggc agc 528
345 Gly Ile Lys Ala His Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
346 165 170 175
348 gtc cag ctg gcc gac cac tac cag cag aac acc ccc atc ggc gac ggc 576
349 Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
350 180 185 190
352 ccc gtc ctg ctg ccc gac aac cac tac ctg agc acc cag tcc gcc ctg 624
353 Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
354 195 200 205
356 agc aaa gac ccc aac gag aag cgc gat cac atg gtc ctg ctg gag ttc 672
357 Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
358 210 215 220
360 gtc acc gcc gcc ggc atc act ctg ggc atg gac gag ctg tac aag taa 720
361 Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
362 225 230 235
365 <210> SEQ ID NO: 6
366 <211> LENGTH: 239
367 <212> TYPE: PRT
368 <213> ORGANISM: Aequorea victoria
370 <400> SEQUENCE: 6

```

## VERIFICATION SUMMARY

DATE: 06/12/2001

PATENT APPLICATION: US/09/866,538

TIME: 15:13:11

Input Set : A:\REGEN1530-2.ST25.txt

Output Set: N:\CRF3\06122001\I866538.raw

L:11 M:270 C: Current Application Number differs. Replaced Current Application Number  
L:855 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 13  
L:867 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 14  
L:879 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 15  
L:891 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 16  
L:903 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 17  
L:915 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 18  
L:927 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 19  
L:939 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 20  
L:951 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 21  
L:970 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 22  
L:982 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 23  
L:994 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 24  
L:1006 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 25  
L:1018 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 26  
L:1034 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 27  
L:1048 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 28  
L:1062 M:220 C: Keyword misspelled or invalid format. <213> ORGANISM for SEQ ID= 29